Yuting Wei

Statistics & Data Science
433 Academic Research Building
The Wharton School, University of Pennsylvania

Email: ytwei@wharton.upenn.edu
Homepage: https://yutingwei.github.io/

Employment

University of Pennsylvania
Assistant Professor, Department of Statistics and Data Science, The Wharton School, 07/2021 - present

Carnegie Mellon University
Assistant Professor, Department of Statistics and Data Science, 08/2019 - 06/2021

Stanford University
Stein Fellow/Lecturer, Department of Statistics, 09/2018 - 07/2019

Education

University of California, Berkeley Ph.D. in Statistics, 2013 - 2018

- Thesis: A geometric perspective on some topics in statistical learning.
- Advisor: Martin Wainwright, Adityanand Guntuboyina

Peking University Bachelor of Science (B.S.) Applied Mathematics, 2009 - 2013

- Advisor: Jinzhu Jia, Peking University

Peking University Bachelor of Arts (B.A.) Economics, 2010 - 2013

Research interests

- High-dimensional and nonparametric statistics
- Statistical machine learning
- Reinforcement learning
- Applications in bioinformatics and statistical genomics

Short-term appointments

Simons Institute for the Theory of Computing, University of California, Berkeley

Visiting scientist, 08/2021 - 12/2021

Institute for Mathematical Research (FIM), ETH Zürich

Visiting researcher, 09/2015 - 12/2015

Department of Mathematics, City University of Hong Kong

Visiting researcher, 08/2012 - 11/2012

Honors & awards

INFORMS George Nicholson Award, Finalist, 2021

Stein Fellowship, Stanford University, 2019

Erich L. Lehmann Citation, University of California, Berkeley, 2018

Oberwolfach Leibniz Graduate Student Grant, 2017

Berkeley Graduate Student Travel Grant, 2017

NeurIPS (Neural Information Processing Systems) Travel Award, 2017

Outstanding Graduates of Peking University, 2013

Outstanding Winner of Annual Peking University Research Competition (top 10/5000), 2013

Best Overall Winner in 2nd Machine Learning Competition in Immunology, 2012

President Research Fund, Peking University, 2011-2012

Journal articles & preprints

J1. Yue Li, **Yuting Wei**, "Minimum ℓ_1 interpolators: Precise asymptotics and multiple descent," arxiv:2110.09502, 2021.

J2. Shicong Cen, **Yuting Wei**, Yuejie Chi, "Fast Policy Extragradient Methods for Competitive Games with Entropy Regularization," arXiv:2105.15186, 2021.

J3. Gen Li, Changxiao Cai, Yuxin Chen, Yuantao Gu, **Yuting Wei**, Yuejie Chi, "Is Q-Learning Minimax Optimal? A Tight Sample Complexity Analysis," arXiv:2102.06548, 2021.

J4. Michael Celentano, Andrea Montanari, **Yuting Wei** (alphabetical order), "The Lasso with General Gaussian Designs with Applications to Hypothesis Testing," arXiv:2007.13716, 2020.

J5. Gen Li, **Yuting Wei**, Yuejie Chi, Yuantao Gu, Yuxin Chen, "Softmax Policy Gradient Methods Can Take Exponential Time to Converge," arXiv:2102.11270, 2021.

J6. Gen Li, **Yuting Wei**, Yuejie Chi, Yuantao Gu, and Yuxin Chen, "Breaking the Sample Size Barrier in Model-Based Reinforcement Learning with a Generative Model," arXiv:2005.12900, 2020.

J7. Zhimei Ren, **Yuting Wei**, Emmanuel J. Candès, "Derandomizing Knockoffs," arxiv:2012.02717, *Journal of the American Statistical Association (JASA)*, 2021.

J8. Minshi Peng, Yue Li, Brie Wamsley, **Yuting Wei**, Kathryn Roeder, "Integration and transfer learning of single-cell transcriptomes via cFIT," *Proceedings of the National Academy of Sciences (PNAS)*, vol. 118, no. 10, 2021.

J9. Minshi Peng, Brie Wamsley, Andrew Elkins, Daniel M Geschwind, **Yuting Wei**, Kathryn Roeder, "Cell Type Hierarchy Reconstruction via Reconciliation of Multi-resolution Cluster Tree," *Nucleic Acids Research*, 2021.

J10. Gen Li, **Yuting Wei**, Yuejie Chi, Yuantao Gu, and Yuxin Chen, "Sample Complexity of Asynchronous Q-Learning: Sharper Analysis and Variance Reduction," arXiv:2006.03041, *IEEE Transactions on Information Theory*, 2021.

J11. Shicong Cen, Chen Cheng, Yuxin Chen, **Yuting Wei** and Yuejie Chi, "Fast Global Convergence of Natural Policy Gradient Methods with Entropy Regularization," arXiv:2007.06558, *Operations Research*, 2021. **INFORMS George Nicholson Paper Award Finalist**.

J12. Chen Cheng, **Yuting Wei**, and Yuxin Chen, "Tackling Small Eigen-gaps: Fine-Grained Eigenvector Estimation and Inference under Heteroscedastic Noise," arXiv:2001.04620, *IEEE Transactions on Information Theory*, 2021.

J13. **Yuting Wei** and Martin Wainwright, "The Local Geometry of Testing in Ellipses: Tight Control via Localized Kolmogorov Widths," *IEEE Transactions on Information Theory*, vol. 66, no. 8, pp. 5110-5129, 2020.

J14. **Yuting Wei**, Billy Fang, and Martin Wainwright, "From Gauss to Kolmogorov: Localized Measures of Complexity for Ellipses," *Electronic Journal of Statistics,* vol. 14, no. 2, pp. 2988-3031, 2020.

J15. **Yuting Wei**, Martin Wainwright and Adityanand Guntuboyina, "The Geometry of Hypothesis Testing over Convex Cones: Generalized Likelihood Tests and Minimax Radii," *The Annals of Statistics,* vol. 47, no. 2, pp. 994-1024, 2019.

J16. **Yuting Wei***, Fanny Yang* and Martin Wainwright, "Early Stopping for Kernel Boosting Algorithms: A General Analysis with Localized Complexities," *IEEE Transactions on Information Theory*, vol. 65, no. 10, pp. 6685-6703, 2019.

J17. Tony Cai, Adityanand Guntuboyina and **Yuting Wei** (alphabetical order), "Adaptive Estimation of Planar Convex Sets," *The Annals of Statistics,* vol. 46, no. 3, pp. 1018-1049, 2018.

J18. Wen-Jun Shen*, **Yuting Wei***, Xin Guo*, Stephen Smale, Hau-San Wong and Shuaicheng Li, "MHC Binding Prediction with KernelRLSpan and Its Variations," *Journal of immunological methods,* vol. 406, pp. 10-20, 2014.

Conference papers

C1. Shicong Cen, **Yuting Wei**, Yuejie Chi, "Fast Policy Extragradient Methods for Competitive Games with Entropy Regularization," *Neural Information Processing Systems (NeurIPS)*, 2021.

C2. Gen Li, Yuxin Chen, Yuejie Chi, Yuantao Gu, **Yuting Wei**, "Sample-Efficient Reinforcement Learning is Feasible for Linearly Realizable MDPs with Limited Revisiting," *Neural Information Processing Systems (NeurIPS)*, 2021.

C3. Gen Li, **Yuting Wei**, Yuejie Chi, Yuantao Gu, and Yuxin Chen, "Softmax Policy Gradient Methods Can Take Exponential Time to Converge," *Conference on Learning Theory (COLT)*, 2021.

C4. Pratik Patil, **Yuting Wei**, Alessandro Rinaldo, Ryan Tibshirani, "Uniform consistency of cross-validation estimators for high-dimensional ridge regression," *International Conference on Artificial Intelligence and Statistics (AISTATS)*, oral presentation (10.5% of accepted papers), 2021.

C5. Gen Li, Changxiao Cai, Yuxin Chen, Yuantao Gu, **Yuting Wei**, Yuejie Chi, "Tightening the Dependence on Horizon in the Sample Complexity of Q-Learning ," *International Conference on Machine Learning (ICML)*, 2021.

C6. Jingyan Wang, Ivan Stelmakh, **Yuting Wei**, Nihar B. Shah, "Debiasing Evaluations That are Biased by Evaluations," *AAAI Conference on Artificial Intelligence*, 2021.

C7. Yue Li, Ilmum Kim, **Yuting Wei**, "Randomized Tests for High-Dimensional Regression: A More Efficient and Powerful Solution," *Neural Information Processing Systems (NeurIPS)*, 2020

C8. Gen Li, **Yuting Wei**, Yuejie Chi, Yuantao Gu, and Yuxin Chen, "Breaking the Sample Size Barrier in Model-Based Reinforcement Learning with a Generative Model," *Neural Information Processing Systems (NeurIPS)*, 2020.

C9. Gen Li, **Yuting Wei**, Yuejie Chi, Yuantao Gu, and Yuxin Chen, "Sample Complexity of Asynchronous Q-Learning: Sharper Analysis and Variance Reduction," *Neural Information Processing Systems (NeurIPS)*, 2020.

C10. Chen Dan, **Yuting Wei** and Pradeep Ravikumar, "Sharp Statistical Guarantees for Adversarially RobustGaussian Classification," *International Conference on Machine Learning (ICML)*, 2020.

C11. **Yuting Wei**, Fanny Yang and Martin Wainwright, "Early Stopping for Kernel Boosting Algorithms: A General Analysis with Localized Complexities," *Neural Information Processing Systems (NeurIPS)*, spotlight presentation, 2017.

C12. **Yuting Wei** and Martin Wainwright, "Sharp Minimax Bounds for Testing Discrete Monotone Distributions," *International Symposium on Information Theory (ISIT),* July 2016.

Short courses and tutorials

S1. "Statistical and Algorithmic Foundations of Reinforcement Learning," *ICSA 2021 Symposium Short Course*, cotaught with Yuejie Chi, Yuxin Chen and Zhengyuan Zhou, Sep. 2021.

Invited talks

T1. "Breaking the Sample Size Barrier in Reinforcement Learning," *Stochastics and Statistics Seminar, MIT*, Oct. 2021.

T2. "Bridge over troubled water: Lower bounds and fast convergence for policy optimization," *Informs Annual Meeting*, Oct. 2021.

T3. "Sample complexity of Q-learning: Sharper analysis and minimax optimality," *ICSA 2021 Symposium*, Sep. 2021.

T4. "Breaking the Sample Size Barrier in Reinforcement Learning," *Statistics Seminar, Duke University*, Sep. 2021.

T5. "Breaking the Sample Size Barrier in Reinforcement Learning," *Joint Statistical Meetings (JSM)*, Aug. 2021.

T6. "Softmax Policy Gradient Methods Can Take Exponential Time to Converge," *Conference on Learning Theory (COLT)*, Aug. 2021.

T7. "Statistical Inference over Convex Cones," *SIAM Conference on Optimization*, July. 2021.

T8. "Modern Statistical Perspectives in Reinforcement Learning and Early Stopping," *Statistics Seminar, Harvard University*, Feb. 2021.

T9. "Breaking the Sample Size Barrier in Reinforcement Learning," *Statistics Seminar, Wharton Statistics Seminar, University of Pennsylvania*, Feb. 2021.

T10. "Modern Statistical Perspectives in Reinforcement Learning and Early Stopping," *Statistics Seminar, Department of Data Sciences and Operations, University of Southern California*, Feb. 2021.

T11. "Breaking the Curse of Dimensionality: From Sparse Regression to Kernel Boosting," *Statistics Seminar, Stern School of Bussiness, New York University*, Jan. 2021.

T12. "Breaking the Sample Size Barrier in Reinforcement Learning," *Statistics Seminar, Yale University*, Jan. 2021.

T13. "Breaking the Sample Size Barrier in Statistical Inference and Reinforcement Learning," *Wilks Statistics Seminar, ORFE, Princeton University*, Dec. 2020.

T14. "Breaking the Sample Size Barrier in Statistical Inference and Reinforcement Learning," *Statistics Seminar, Rutgers University*, Dec. 2020.

T15. "Breaking the sample size barrier in model-based reinforcement learning with a generative model," *Informs Annual Meeting*, Nov. 2020.

T16. "Breaking the sample size barrier in model-based reinforcement learning with a generative model," *Richard Samworth's group, Cambridge University*, Nov. 2020.

T17. "Reliable hypothesis testing paradigms in high dimensions," *Columbia University, Statistics Seminar*, Oct. 2020.

T18. "Reliable hypothesis testing paradigms in high dimensions," *Young Data Science Researcher Seminar Zurich, ETH*, Oct. 2020.

T19. "Reliable hypothesis testing paradigms in high dimensions" *Michigan State University, Statistics Seminar*, Oct. 2020.

T20. "The Lasso with general Gaussian design with application to hypothesis testing," *Joint Statistical Meeting*, Aug. 2020.

T21. "Breaking the sample size barrier in model-based reinforcement learning with a generative model," *TBSI*

Workshop on Learning Theory, Shenzhen, Jul. 2020.

T22. "Understanding the distribution of the Lasso and its applications," *STATML Group, Carnegie Mellon University,* Mar. 2020.

T23. "A geometric perspective on hypothesis testing," *ICSA International Conference, Hangzhou, China,* Dec. 2019.

T24. "Towards a better understanding of the regularization in kernel learning," *TBSI Workshop on Learning Theory*, Shenzhen, Dec. 2019.

T25. "Towards a better understanding of the regularization in kernel learning," *Statistics Seminar, Cambridge University,* Nov. 2018.

T26. "Towards a better understanding of the regularization in kernel learning," *Big Data and Computational Social Science Lecture, University of British Columbia,* Mar. 2019.

T27. "Early stopping for gradient type algorithms," *Computational and Methodological Statistics, University of Pisa, Italy,* Dec. 2018.

T28. "The geometry of hypothesis testing over convex cones," *52nd Annual Conference on Information Systems and Sciences (CISS) conference, Princeton,* Mar. 2018.

T29. "Geometric analysis of hypothesis testing and early stopping for boosting," *Statistics Seminar, Carnegie Mellon University,* 2018.

T30. "Geometric analysis of hypothesis testing and early stopping for boosting," *Statistics Seminar, Stanford University,* 2018.

T31. "Geometric analysis of hypothesis testing and early stopping for boosting," *Statistics Seminar, University of Michigan,* 2018.

T32. "Geometric analysis of hypothesis testing and early stopping for boosting," *Statistics Seminar, University of Pennsylvania,* 2018.

T33. "Shape-constrained methods: Inference, applications, and practice," *Banff International Research Station for Mathematical Innovation and Discovery (BIRS), Canada,* 2018.

T34. "Early stopping for kernel boosting algorithms," *Neural Information Processing Systems (NIPS) conference, Long Beach,* 2017.

T35. Student talk for Oberwolfach Workshop "Statistical Recovery of Discrete, Geometric and Invariant Structures", 2017.

T36. "Sharp minimax bounds for testing discrete monotone distributions," *International Symposium on Information Theory (ISIT) conference, Barcelona,* 2016.

T37. "Adaptive estimation of planar convex sets," *Berkeley Statistics Annual Research Symposium (BSTARS),* 2016.

T38. "Sharp minimax bounds for testing discrete monotone distributions." *Stanford-Berkeley Joint Colloquium,* 2016.

Ph. D. supervision

1. Yue Li (co-advised with Kathryn Roeder)
2. Minshi Peng (co-advised with Kathryn Roeder)

Doctoral committees

Chen Dan (CMU, Computer Science), Pratik Patil (CMU, Statistics & Machine Learning), Yufei Yi (CMU, Statistics)

Teaching

Carnegie Mellon University:

- 36-747: Mathematics for High Dimensional Data: A statistical viewpoint, Spring 2021
- 36-748: Mathematics for High Dimensional Data: An optimization viewpoint, Spring 2021
- 36-225: Introduction to Probability, Fall 2020
- 36-741: Statistics meets Optimization: Iterative sketching methods, Fall 2019
- 36-742: Statistics meets Optimization: Approximate message passing algorithms, Fall 2019

Stanford University:

- Statistics 314: Advanced Statistical Theory, Spring 2019
- Statistics 206: Applied Multivariate Statistical Analysis, Winter 2018

Berkeley (Graduate Student Instructor):

- STAT 210A: Theoretical Statistics, Fall 2014
- STAT 135: Concepts of Statistics, Spring 2014
- STAT 134: Concepts of Probability, Summer 2014

Selected professional service

Reviewer for Journals: the Annals of Statistics, Journal of Machine Learning Research, IEEE Transactions on Information Theory, Statistica Sinica, Electronic Journal of Statistics, Journal of the Korean Statistical Society

Reviewer for Conferences: International Conference on Artificial Intelligence and Statistics (AISTATS), IEEE International Symposium on Information Theory (ISIT), Conference on Neural Information Processing Systems (NeurIPS), International Conference on Machine Learning (ICML), Annual Conference on Learning Theory (COLT), Asian Conference on Machine Learning (ACML)

University and department service

Statistics Department Seminar Co-organizer at CMU, 2020 - 2021

Faculty Senate Member at CMU, 2020 - 2021

Statistics Department Seminar Co-organizer at CMU, 2019 - 2020

Faculty Senate Member at CMU, 2019 - 2020

Statistics Department Seminar Co-organizer at Stanford, 2018 - 2019